

# Aplicación para el análisis y tendencias poblacionales del COVID-19 en México por medio de minería de datos

Joaquín Chávez-Martínez  
Unidad Académica de Ingeniería  
Eléctrica  
Univerisdad Autónoma de Zacatecas  
Zacatecas, México  
joaquintoto15@gmail.com

Francisco Muro-Bueno  
Unidad Académica de Ingeniería  
Eléctrica  
Univerisdad Autónoma de Zacatecas  
Zacatecas, México  
francisco.muro4f@gmail.com

Aldonso Becerra-Sánchez  
Unidad Académica de Ingeniería  
Eléctrica  
Univerisdad Autónoma de Zacatecas  
Zacatecas, México  
a7donso@uaz.edu.mx

Gustavo Zepeda-Valles  
Unidad Académica de Ingeniería  
Eléctrica  
Univerisdad Autónoma de Zacatecas  
Zacatecas, México  
gzepeda@uaz.edu.mx

**Abstract**—La reciente pandemia del COVID-19 ha traído consigo nuevos problemas a los que el mundo moderno no se había enfrentado, desde temas de salud, hasta consecuencias económicas, sociales y políticas. En el tema de salud ya es usual que se implementen diferentes técnicas de Machine Learning para ayudar a diferentes problemas, como sistemas que ayudan al tratamiento, diagnóstico, entre otros. Existe un dataset que provee información actualizada de los pacientes sospechosos o positivos de COVID-19, dicho dataset oficial es proporcionado por el Gobierno de México; sin embargo, este solo se aprovecha para la creación de gráficas que muestran el avance de la enfermedad. Este artículo tiene como objetivo crear un prototipo de un sistema web que brinde análisis descriptivo e inferencial de la información de este dataset, ayudando a la predicción de la probabilidad de no sobrevivir a la enfermedad, de hospitalización y desarrollo de síntomas graves, como lo puede ser la intubación, utilizando datos categóricos de factores de riesgo. Para ello se utilizaron 4 modelos de redes neuronales, la primera predice la probabilidad de desarrollar síntomas graves con una precisión de 98%; la segunda red predice la probabilidad de fallecimiento con una precisión de 92.4%; la tercera red neuronal predice la probabilidad de hospitalización con una precisión de 85.6%; la última red neuronal predice la probabilidad de no recuperarse a la enfermedad, con base en dos fechas, con una precisión de 92.3%. Al mismo tiempo se emplearon modelos de regresión logística, obteniendo una precisión de 98% en el modelo que predice si se tendrán síntomas severo, un 92.3% en el esquema que predice si fallecerá, mientras que un 85.6% de precisión en el modelo que predice si el paciente será hospitalizado. Los resultados muestran que el prototipo es considerablemente confiable y que podría ser utilizado dentro de un ambiente controlado; así mismo que se puede emplear para ayudar a personal apropiado para catalogar a pacientes y conocer qué medidas pudieran ser tomadas en cuenta acorde a las predicciones, de tal manera que se puede tener mayor o menor concentración en ellos.

**Keywords**—regresión logística, redes neuronales, predicciones de salud, COVID-19, aplicación web.

## I. INTRODUCCIÓN

A finales del año 2019 el virus conocido como SARS-CoV-2 comenzó a esparcirse por todo el mundo. Este hecho dio comienzo a una de las pandemias más peligrosas del mundo; su propagación fue rápida por China, ya que en un mes esta enfermedad, que logra esparcirse velozmente, fue catalogada como “emergencia de salud pública de alcance internacional” [1], [2]. A los 31 días del mes de diciembre 2019, el Ministerio de Salud de China informó a la ON

(Organización Mundial de la Salud) que se tenían 41 pacientes con “neumonía atípica grave”; la mayoría de ellos relacionados con ingesta de comida proveniente de animales “exóticos” (de mar o tierra) en la ciudad de Wuhan, provincia de Hubei en China [3].

Al final de cuentas, durante poco más de un año y medio el tema de interés por mucho ha sido la salud y estar informados acerca de este nuevo virus. Pero ¿qué es el COVID-19? Este virus también conocido como SARS-CoV-2 es un nuevo beta-coronavirus de la familia Coronaviridae, llamado así por su forma esférica rodeado por múltiples espículas (glicoproteínas-S) que le dan un aspecto parecido al de una corona [4], [5]. Este virus es transmitido en su mayoría por los aerosoles que las personas exhalan al momento de hablar y respirar. Su índice de contagio es alto ( $R_0$  = número de reproducibilidad hasta 5), lo que hace que una sola persona infectada pueda contagiar de 2 a 5 personas; donde el periodo de incubación de este virus es de 1 a 14 días, con una media de 5 a 6 días [6], [7].

Con el pasar de los días, la pandemia ha cobrado demasiadas víctimas en México, hablamos de 296,197 muertes por COVID-19 aproximadamente al día 12 de octubre del 2021 [8].

Esta pandemia creó una serie de conflictos que no se habían enfrentado antes, la desinformación siendo uno de estos. La OMS señala que el comienzo de la pandemia fue acompañado con una masiva pandemia de desinformación (del inglés ‘Infodemic’), esto haciendo que la información disponible fuera poco confiable para la gente cuando lo necesitaran [9]. Al principio de esta pandemia también se hablaba de cómo los países no tenían manera en la que se pudieran comunicar efectivamente con su población, esto para mantenerlos informados de la situación; este comunicado fue emitido a principios de febrero por la OMS. Actualmente México cuenta con una página oficial respecto a COVID-19 en la cual muestran los positivos estimados, defunciones estimadas, activos estimados a nivel nacional, entre otros datos estadísticos, junto con un semáforo epidemiológico que muestra la densidad de personas contagiadas por COVID-19 según el estado [10]. Dentro de esta página no se puede encontrar ninguna herramienta que pueda ayudar a saber, si con base a los factores de riesgo de una persona, esta pueda ser hospitalizada, o padecer de síntomas graves o incluso leer. Pero ¿es posible predecir si a una persona le pueden

sucedan estos acontecimientos con base en las muestras que existen?

En este artículo se utilizan técnicas de Machine Learning sobre este conjunto de datos oficial para proveer una herramienta que pueda ayudar al usuario a visualizar información procesada, las cuales predicen la probabilidad de padecer síntomas graves, ser hospitalizado o morir, todo esto con base en ciertos factores de riesgo y desde una aplicación web. En esta investigación se entrenan redes neuronales y modelos de regresión logística para lograr predecir de manera satisfactoria este tipo de información, además de integrarlo como una aplicación web. En este artículo se explica el tipo de redes neuronales que se utilizaron, cómo se implementó la regresión logística, así como a qué tipo de problemas se enfrenta de manera teórica; también se describe la normalización del dataset usado, incluyendo la integración de todo esto en una aplicación web desarrollada con Django [11]. La precisión obtenida de las redes neuronales fue de 85.6% en el esquema que predice el tipo de paciente (es decir si es hospitalizado o ambulatorio), 98% en el modelo que predice si se pueden tener síntomas graves como intubación, y 92.4% en la red que predice la probabilidad de morir. Todo esto con base a factores de riesgo, como diabetes, asma, EPOC (enfermedad pulmonar obstructiva crónica), entre otros; además de la edad, la cual se ha visto como un factor importante a considerar en este tópico en específico. Por último, se obtuvo un 92.3% de precisión en la red neuronal que predice la probabilidad de fallecimiento basándose en la fecha que el paciente presentó síntomas contra la fecha que fue ingresado a la unidad médica. Como cálculos adicionales se obtuvieron el recall, la cual dio como resultado 0.902 para la red neuronal que predice el tipo de paciente, mientras que su especificidad dio como resultado 0.904. Para la red neuronal que calcula la probabilidad de ser intubado, debido a las enfermedades del paciente, su recall dio como resultado 0.907, mientras que su especificidad salió de 0.866. Para la red neuronal que predice el posible fallecimiento del paciente, se obtuvo un 0.957 para su recall, mientras que 0.929 para su especificidad. Por último, la red neuronal que predice el posible fallecimiento del paciente según la demora en la que presentó síntomas y la fecha en la que fue ingresado al hospital, obtuvo una precisión de 92.29%, un recall de 0.775, así como una especificidad de 0.745.

Por otro lado, estas mismas predicciones fueron hechas con el algoritmo de regresión logística. La precisión de la regresión logística encargada de predecir el tipo de paciente fue de 85.6%, con un F1-Score de 0.92, además de un recall de 0.97. El modelo que calcula la probabilidad de que algún paciente fuera intubado tiene una precisión de 98%, un F1-Score de 0.99 y un recall prácticamente de 1.0. Por último, la regresión que se encarga de calcular la posible defunción de un paciente tuvo una precisión de 92.3%, un recall de 0.99 y un F1-Score fue de 0.96.

Estas fueron las dos técnicas de Machine Learning utilizadas en esta investigación. Como objetivo adicional, se creó una red neuronal que predice la probabilidad de morir basándose en dos fechas que se encontraban en el dataset. Una de las fechas hace referencia a cuándo el paciente comenzó a presentar síntomas, mientras que la otra a cuándo fue ingresado el paciente a la UM (unidad médica); dando así una precisión de 92.3%. Partiendo de esto, el hecho del tiempo que tarda un paciente en ser ingresado a la unidad médica, una vez teniendo síntomas, puede ser clave para la evolución de la

enfermedad. Es así que existen ciertos aspectos a tomar en cuenta en esta red neuronal y su veracidad, que serán explicados en este artículo.

El artículo está estructurado de la siguiente manera, en la sección II se podrán encontrar trabajos relacionados al tema según la revisión de la literatura. En la sección III se describirá el diseño y desarrollo de la aplicación, así como una explicación de las técnicas utilizadas y el dataset. En la sección IV se explican los problemas por los que se pasó, las soluciones que se les dio, los resultados obtenidos y cómo podemos mejorarlos, así como las limitaciones. Por último, en la sección V se plasman las conclusiones a las que se llegaron.

## II. REVISION DE LITERATURA

A principios de la pandemia, varios autores comenzaron a utilizar conjuntos de datos de diferentes procedencias, los cuales disponen de información médica de los pacientes diagnosticados o sospechosos de tener COVID-19. Estos han tenido como propósito aplicar técnicas de Machine Learning y lograr crear herramientas útiles para diversos temas relacionados con el COVID-19.

En la Universidad de Madrid comenzaron a realizar análisis de la mortalidad en pacientes con COVID-19, utilizando regresión logística, redes Bayesianas, árboles de decisión y bosques aleatorios, además utilizaron Biclustering para el análisis. Estos trabajos concluyeron que la regresión logística era altamente confiable, utilizando como variables la edad, saturación de oxígeno, institución residencial, y deterioro oncológico del paciente, obteniendo un ABC (Area Bajo la Curva) de 0.89, un recall de 0.80 y una especificidad de 0.83. Los bosques aleatorios también fueron altamente confiables, en ellos utilizaron como variables independientes la saturación de oxígeno, ritmo cardíaco, institución residencial y temperatura, todo esto medido en las salas de emergencia; obteniendo un AUC de 0.87, un recall de 0.82 y una especificidad de 0.75. Estos dos modelos fueron los más confiables de los diferentes enfoques utilizados [12].

Otros trabajos hacen un análisis epidemiológico con un conjunto de datos del repositorio oficial de la Universidad de Johns Hopkins, utilizando para ellos RVA (regresión de vectores de apoyo), RNP (red neuronal profunda), RMCP (redes de memoria a corto plazo) y RP (regresión polinómica). Todas las técnicas fueron usadas para encontrar una predicción de cómo avanzaría la enfermedad en la sociedad, RP brindó el menor error cuadrático medio a comparación de los otros enfoques. Esto dio como conclusión que si el esparcimiento de la predicción era correcta, llevaría a una gran pérdida de vidas, ya que presentaba un crecimiento exponencial a nivel mundial; sin embargo, dicho esparcimiento puede ser reducido si se siguen las medidas de cuarentena y distanciamiento social [13].

Miembros de la IEEE crearon un artículo comentando las mejores opciones para atacar diferentes problemas que nacieron del COVID-19, esto con el fin de ayudar a investigadores que buscan crear una herramienta que pueda apoyar a los profesionales de la salud en diferentes tareas. Por ejemplo, usar Extreme Learning Machine para encontrar el uso adecuado de drogas, basados en una serie específica de variables, y poder ayudar a reducir la carga de los cardiólogos; saber si el paciente irá a UCI (Unidad de

Cuidados Intensivos), o si incluso será hospitalizado, cuanto tiempo padecerá los síntomas, entre otros. En este sentido, también usar RMCP para predecir el uso reflexivo de los marcadores biológicos con una atención diferenciada, incluyendo la estimación de la afectación cardíaca causada por el COVID-19. Emplear RNN (redes neuronales recurrentes) para una predicción de la propagación del COVID-19. Adicionalmente, utilizar RAG (red adversarial generativa) para la visualización de detección de nuevos coronavirus humanos. Todo esto como una herramienta de ayuda para los investigadores [14].

Otros investigadores crearon una RNP para predecir la probabilidad de que una persona termine en la UCI. Se entrenó con diferentes datos, desde la edad y la etnia, hasta variables más clínicas como fiebre, toz, fatiga, entre otros, llegando así a tener una ABC de 0.780, lo cual es una salida aceptable. Eso se buscó con el propósito de dar una herramienta de primera línea para poder tratar a los pacientes de manera diferente con base en sus factores de riesgo, ya que el tiempo es un factor importante [15].

En Reino Unido, crearon un sistema de puntuación del riesgo de mortalidad que tiene un paciente en el punto de ingreso utilizando una RNA. Esto fue realizado basándose en una muestra de 298 pacientes admitidos en un hospital, donde el objetivo fue poder tratarlos y tener diferentes cuidados, analizando demografía, comorbilidades, historial de tabaquismo y los síntomas que presenta el paciente. Como resultado se obtuvieron predicciones con un 86.5% de precisión, un recall de 87.5% y una especificidad de 85.94%, demostrando así la gran utilidad que tiene el uso de Deep Learning en el tema del COVID-19 [16].

En otro artículo se utilizaron algoritmos de bosques aleatorios para hacer predicciones acerca del COVID-19, creando su propio conjunto de datos con combinación de datos de la Organización Mundial de la Salud y la Universidad de John Hopkins, el cual fue pre-procesado. Con un resultado de F1-score de 0.86 en dicho conjunto de datos, se descubrió que provee predicciones precisas, incluso en conjuntos de datos desequilibrados. Concluyendo así que la tasa de muerte fue mayor entre pacientes nativos de Wuhan comparado con no nativos. También los hombres tuvieron mayor tasa de muerte comparado con las mujeres; donde la mayoría de los pacientes afectados se encontraron entre los 20 y los 70 años de edad [17].

Se puede notar que la idea de utilizar técnicas de inteligencia artificial se ha venido dando con grandes resultados en esta área de la salud. La mayoría de los artículos que hablan de esto fueron publicados entre junio y septiembre de 2020, en donde se incluyen conjuntos de datos que datan de febrero a mayo aproximadamente, ya que fue cuando los primeros conjuntos de datos comenzaron a estar disponibles. No obstante, siempre es útil disponer de herramientas que ayuden al procesamiento automático y semi-automático de la información para el usuario final, lo cual es uno de los objetivos principales de la presente preputa.

### III. METODOLOGIA

En este apartado se podrán revisar que métodos fueron empleados para seleccionar los datos que fueron utilizados para el entrenamiento de las redes neuronales y la regresión

logística. Además, se revisará la arquitectura del sistema y qué tipo de tecnologías fueron utilizadas para su desarrollo.

#### A. Metodología de prototipo

Para el desarrollo de este sistema se optó por una metodología ágil conocida como Prototyping. Esta metodología se define como un método de desarrollo de sistemas en el que un prototipo (una aproximación temprana de un sistema final o producto) es construido, probado y luego reinventado según sea necesario. El proceso continuo hasta que finalmente se logre un prototipo aceptable a partir del cual el sistema o producto completo puede ser desarrollado [18]. Este modelo de prototipo permite obtener retroalimentación de algunos de los interesados en el sistema en etapas tempranas, haciendo que funcionalidades del sistema sean descartadas o suprimidas, mientras que nuevas funcionalidades y necesidades sean agregadas conforme van siendo requeridas. El prototipo es una versión preliminar, intencionalmente incompleta o reducida de un sistema [19]. El uso de prototipos es una herramienta útil para aplicarse en casi todas las actividades de diseño y creación de software. Para el caso de las cuestiones tácticas del presente trabajo, el prototipo diseñado se basará en las especificaciones de los requerimientos bajo los supuestos de:

- a) Definición de los objetivos del prototipo.
- b) Validación de los requerimientos a través de la funcionalidad del producto.
- c) Diseñar y desarrollar el prototipo.
- d) Evaluar el funcionamiento del prototipo.

El enfoque Prototyping sigue un proceso que se repite en cada iteración [20], [21] (ver Fig.1). Durante el establecimiento de los objetivos del prototipo, el fin es seleccionar una de las funcionalidades a desarrollar y definir hasta donde llegaría el prototipo en etapa de evaluación. Esta tarea requiere de la selección cuidadosa de características que requieran ser evaluadas por el usuario final para su aprobación o desecho. Cada uno de estos pasos nos permite mejorar el sistema, incluso antes de que esté finalizado.

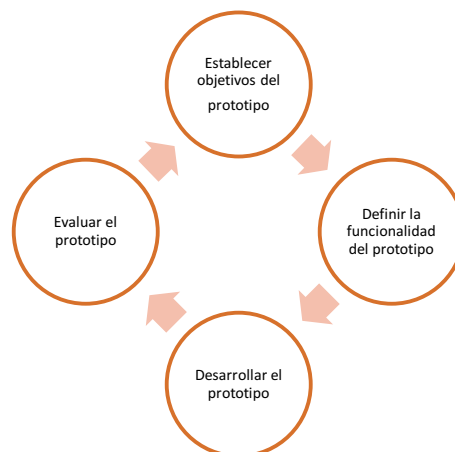


Fig. 1. Ciclo de vida de prototipo

#### B. Definición de los objetivos del prototipo

En este prototipo se tuvieron en mente los siguientes objetivos:

- OBJ-1. Crear una aplicación que facilite el análisis automático del dataset de Gobierno de México, brindando información integral de los datos.
- OBJ-2. Informar, con base en conjuntos de datos oficiales de la Federación, estadísticas básicas de la actual pandemia del COVID-19, como edades, estados con mayor densidad de casos, entre otros.
- OBJ-3. Informar, con base en los factores de riesgo, edad y otras características, una predicción fiable de si el paciente diagnosticado con COVID-19 positivo es más propenso a ser hospitalizado; o si será más bien un paciente de tipo ambulatorio.
- OBJ-4. Informar, con base en los factores de riesgo, edad y otras características, una predicción fiable de si el paciente con diagnóstico COVID-19 es más propenso a desarrollar síntomas severos como la intubación.
- OBJ-5. Informar, con base en los factores de riesgo, edad y otras características, una predicción fiable de si el paciente con diagnóstico COVID-19 es propenso a fallecer.
- OBJ-6. Informar, con base en la fecha que se presentaron los síntomas y la fecha en la que fue ingresado el paciente a la unidad médica, la probabilidad de que este fallezca.

C. Validación de requerimientos a través de la funcionalidad

El propósito de esta herramienta es ayudar a usuarios del sector salud y al público en general a tener una predicción de cómo la enfermedad puede avanzar según diferentes variables. Se espera que esta aplicación de un apoyo al momento de tener el resultado positivo a la enfermedad del COVID-19, para una mejor toma de decisiones alrededor de estos. De igual manera se espera que este prototipo sea capaz de predecir diferentes situaciones con base en las características que muestre un paciente que sea positivo a COVID-19, además de proveer información epidémica de las personas contagiadas en el país.

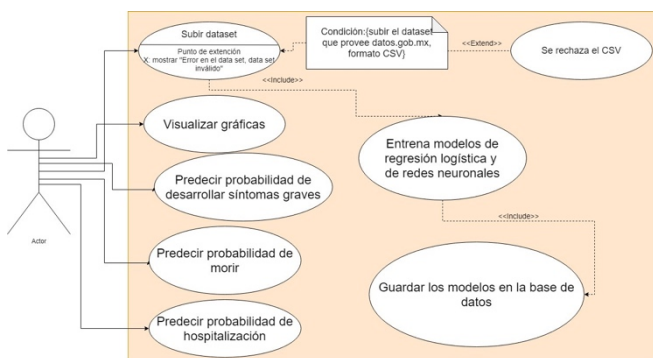


Fig. 2. Diagrama de Casos de Uso

En la Fig. 2. se puede observar que el actor o usuario final puede intervenir en diversos casos de uso:

- “Subir dataset”, permite al usuario subir un dataset más actualizado, obtenido de la página oficial de federación, ya que solo funcionará con ese dataset.
- Si el CSV (Comma Separated Values, tipo de archivo separado por comas) se sube sin problemas y cumple con los requerimientos, entonces se

“Entrenan los modelos de regresión logística y redes neuronales”.

- Después de que se entrenan los modelos, el sistema procede a “Guardar los modelos en la base de datos”, esto para no tener que reentrenar los modelos de manera innecesaria.
- Si el CSV tiene algún problema o situación, entonces “Se rechaza el CSV”.
- “Visualizar Gráficas”, el sistema muestra gráficas de los porcentajes que se encuentran en el dataset; por ejemplo, en género muestra el porcentaje de hombres y de mujeres que hay en el dataset.
- “Predecir probabilidad de desarrollar síntomas graves”, permite al usuario introducir los factores de riesgo que tiene, además de su edad; después el sistema arroja la probabilidad que tiene ese paciente de desarrollar síntomas severos.
- “Predecir probabilidad de morir”, permite al usuario introducir los factores de riesgo que tiene, además de su edad; después el sistema arroja la probabilidad que tiene de no recuperarse.
- “Predecir probabilidad de hospitalización”, el sistema permite hacer lo mismo que en las otras predicciones; después arroja la probabilidad que se tiene de ser hospitalizado por la enfermedad.

D. Diseñar y desarrollar el prototipo.

En este apartado se explica el origen y contenido del dataset de COVID-19. También se describe el flujo de datos esperado para la aplicación web y cómo se seleccionaron las tecnologías utilizadas para su desarrollo.

1) Diseño arquitectónico

Al analizar el problema que se abordaría, y considerando la accesibilidad del programa, se tomó la decisión de hacer el prototipo en una aplicación web. Con esto en mente, las necesidades de poder utilizar un lenguaje de programación que ofreciera las herramientas suficientes para poder cumplir con los objetivos planteados era la meta principal. Al pensar en estadística es normal pensar que el lenguaje llamado R sea una opción viable, aunque no del todo, ya que también es necesario que el lenguaje ofrezca características de desarrollo web. Es así como se llegó a la conclusión de utilizar Python, lenguaje que por excelencia se le conoce por ser versátil, al grado permitir desarrollar programas desde aplicaciones de escritorio, así como aplicaciones web; igualmente, este lenguaje de programación ofrece librerías capaces de facilitar el análisis de datos, como lo son Numpy, Keras, Tensor Flow entre otras [27].

Python brinda frameworks como Django, el cual ofrece la capacidad de desarrollar páginas web utilizando a Python como lenguaje para backend. Django es un framework que utiliza la arquitectura MTV (Modelo Template View), la cual es un concepto similar a la arquitectura MVC (Modelo Vista Controlador), que es conocida por tener 3 características principales.

- M: modelo, parte encargada de almacenar los datos y la persistencia de los mismos; en nuestro modelo se empleó la base de datos sqlite3.
- T: template (Vista), encargada de mostrar todo lo visual relacionado con lo que ve el usuario que interactúa con el sistema; el template es la página HTML como tal, que fue generada con Bootstrap 4.
- V: view (Controlador), es la parte encargada de manejar la lógica del sistema esta se encarga de solicitar los datos al modelo, al igual que llevar los datos a la vista donde serán presentados; en este caso se utilizaron diferentes librerías (ver Fig. 3).

Para Django, en lugar de hacerse llamar Controlador (en el patrón MVC), se denomina View, ya que internamente la lógica dentro del Framework es manejada por diferentes views, que son clases que permiten renderizar un componente con tan solo hacer instancia de él en HTML (HyperText Markup Language). Al igual que el template, este hace referencia a los HTML donde se renderizaran los componentes para la aplicación web [11].

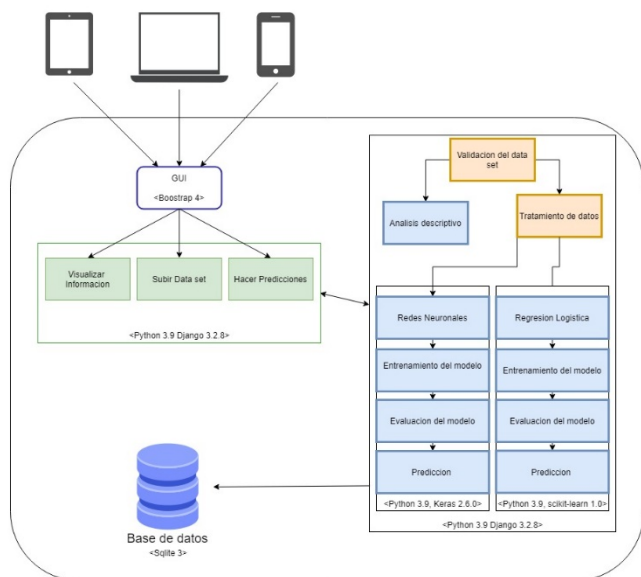


Fig. 3. Diseño de la arquitectura de la aplicación

En la Fig. 3 se puede observar el diagrama de la arquitectura del sistema propuesto, mostrando el modelo MVC que implementa Django. En la parte del controlador, que controla la lógica del sistema, se puede visualizar que lo primero que hace el sistema es validar el dataset; esto para verificar que tenga el formato adecuado, así como las columnas adecuadas. Una vez que se valida el dataset, se hace el tratamiento de datos necesario para los modelos, a la vez que se hace el análisis descriptivo. Luego, cuando entra a las redes neuronales, se entrena el modelo, seguido de una evaluación del mismo, para después estar listo para poder hacer una predicción con base en la información que pueda introducir el usuario. Lo mismo pasa para la regresión logística, y mientras se van haciendo estos pasos, se van guardando en la base de datos los modelos entrenados. En el caso de los análisis descriptivos, se pueden generar gráficas de información general de los datos. A todo esto, el usuario puede tener acceso desde cualquier navegador web.

## 2) Descripción del Dataset

El dataset fue obtenido en la página oficial de datos del Gobierno de México [22], esto nos da la certeza que es un dataset oficial y verídico. El dataset consta de 9,971,374 registros y un total de 40 variables, cada una de las cuales con un tipo de dato diferente (ver TABLA 1).

TABLA I. VARIABLES DEL DATASET Y SUS CARACTERÍSTICAS

Nombre	Tipo de dato	Descripción
Fecha de actualización	Fecha	La base de datos se alimenta diariamente. Esta variable permite identificar la fecha de la última actualización.
Id registro	Cadena	Identificador del caso.
Origen	Entero	Identifica si la información fue sacada por el Sistema de Unidades de Salud Monitoras de Enfermedades Respiratorias o no.
Sector	Entero	Identifica el tipo de institución del Sistema Nacional de Salud que brindó la atención.
Entidad UM (unidad médica)	Entero	Identifica la entidad donde se ubica la unidad médica que brindó la atención.
Sexo	Entero	Identifica al sexo del paciente.
Entidad nacimiento	Entero	Identifica la entidad de nacimiento del paciente.
Entidad residencia	Entero	Identifica la entidad de residencia del paciente.
Municipio residencia	Entero	Identifica el municipio de residencia del paciente.
Tipo de paciente	Entero	Identifica el tipo de atención que recibió el paciente en la unidad. Se denomina como ambulatorio si regresó a su casa o se denomina como hospitalizado si fue ingresado a hospitalización
Fecha de ingreso del paciente	Fecha	Identifica la fecha de ingreso del paciente a la unidad de atención.
Fecha que el paciente comenzó a presentar síntomas	Fecha	Identifica la fecha en que inició la sintomatología del paciente.
Fecha de fallecimiento	Fecha	Identifica la fecha en que el paciente falleció.
Intubado	Entero	Identifica si el paciente requirió de intubación.
Neumonía	Entero	Identifica si al paciente se le diagnosticó con neumonía.
Edad	Entero	Identifica la edad del paciente.
Nacionalidad	Entero	Identifica si el paciente es mexicano o extranjero.
Embarazo	Entero	Identifica si la paciente está embarazada.
Habla lengua indígena	Entero	Identifica si el paciente habla lengua indígena.
Indígena	Entero	Identifica si el paciente se auto-identifica como una persona indígena.
Diabetes	Entero	Identifica si el paciente tiene un diagnóstico de diabetes.
EPOC (enfermedad pulmonar obstructiva crónica)	Entero	Identifica si el paciente tiene un diagnóstico de EPOC.
Asma	Entero	Identifica si el paciente tiene un diagnóstico de asma.
Inmunosupresión	Entero	Identifica si el paciente presenta inmunosupresión.
Hipertensión	Entero	Identifica si el paciente tiene un diagnóstico de hipertensión.



Otra complicación	Entero	Identifica si el paciente tiene diagnóstico de otras enfermedades.
Cardiovascular	Entero	Identifica si el paciente tiene un diagnóstico de enfermedades cardiovasculares.
Obesidad	Entero	Identifica si el paciente tiene diagnóstico de obesidad.
Renal crónica	Entero	Identifica si el paciente tiene diagnóstico de insuficiencia renal crónica.
Tabaquismo	Entero	Identifica si el paciente tiene hábito de tabaquismo.
Otro caso	Entero	Identifica si el paciente tuvo contacto con algún otro caso diagnosticado con SARS CoV-2
Toma muestra laboratorio	Entero	Identifica si al paciente se le tomó muestra de laboratorio.
Resultado laboratorio	Entero	Identifica el resultado del análisis de la muestra reportado por el laboratorio.
Toma muestra antígeno	Entero	Identifica si al paciente se le tomó muestra de antígeno para SARS-CoV-2
Resultado antígeno	Entero	Identifica el resultado del análisis de la muestra de antígeno tomada al paciente
Clasificación final	Entero	Identifica si el paciente es un caso de COVID-19 según el catálogo "CLASIFICACION FINAL".
Migrante	Entero	Identifica si el paciente es una persona migrante.
País de nacionalidad	Entero y cadena	Identifica la nacionalidad del paciente.
País de origen	Entero y cadena	Identifica el país del que partió el paciente rumbo a México.
UCI (Unidad de Cuidados Intensivos)	Entero	Identifica si el paciente requirió ingresar a una Unidad de Cuidados Intensivos.

### 3) Técnicas de Machine Learning

Para las técnicas de predicción se emplearon las redes neuronales artificiales y la regresión logística. Una red neuronal se compone de por lo menos 3 capas, estas incluyen, la capa de entrada, en donde se ingresan las variables independientes; la segunda capa se le llama capa oculta, en donde se pueden tener una capa o varias de ellas, y donde cada capa consta de *n* número de nodos, los cuales tienen comúnmente un peso, un sesgo y una función de activación; al final se tiene la capa de salida, que brinda el resultado predicho [23]. Para este proyecto se utilizó una red neuronal que consta de la capa de entrada, 3 capas ocultas (dos de estas con la función de activación relu), además de la capa de salida (que incluyó la función de activación sigmoidea). Como optimizador se utilizó Adam, un algoritmo de optimización que ayuda a actualizar cada uno de los pesos de la red de manera iterativa usando el dataset de entrenamiento, con una tasa de aprendizaje predeterminada de 0.001 [24]. Se utilizó la entropía cruzada binaria como función de costo, por la cual se manejó la última capa como sigmoidea, que brinda la probabilidad de que ocurra un evento (véase Fig. 4).

Por otro lado, la regresión logística es un modelo de probabilidad que brinda un resultado dada una o varias variables de entrada; la más usada es el modelo binario en el que solo existen dos tipos de resultados (1/0). A pesar del nombre, este es un modelo de clasificación más que de regresión, que incluye un modelo más simple y más eficiente para problemas binarios y lineales, de acuerdo a (1). La regresión logística es una técnica muy utilizada para clasificación en la industria [25], en la que el rango de valores proporcionado por la función sigmoidea oscila entre 0 y 1.

$$F(X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}} \quad (1)$$

Donde  $\beta_0$  y  $\beta_1 x$  son equivalentes al modelo lineal  $y=ax+b$ . Esto básicamente explica cómo la probabilidad de un evento depende de las variables independientes.

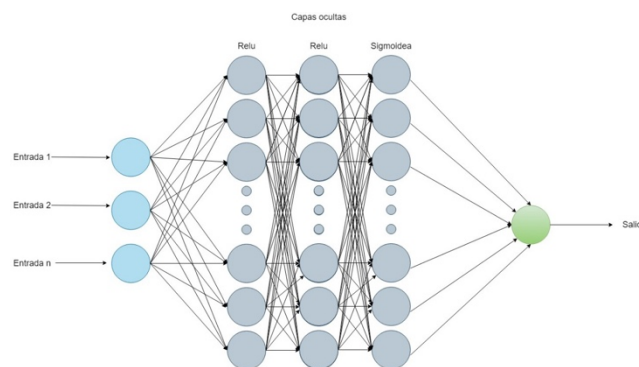


Fig. 4. Modelo de red neuronal empleada

Para que estos modelos pudieran ser entrenados óptimamente es importante normalizar los datos, ya que si se intentan entrenar modelos con números muy grandes y fuera de un parámetro puede comprometer el entrenamiento de los modelos, por eso es importante usar la normalización. En este trabajo se empleó una técnica de normalización simple de acuerdo a (2) [26].

$$X_{\text{normalizada}} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (2)$$

### 4) Desarrollo del prototipo

En este apartado se explica el proceso de desarrollo del prototipo, que incluye el procesamiento del dataset de acuerdo a su utilidad. También se indica cómo se abordaron los objetivos planteados para el desarrollo de la aplicación y cómo fueron abordados cada uno de ellos (ver TABLA II).

TABLA I. VERIFICACIÓN DE CUMPLIMIENTO DE OBJETIVOS

Objetivo	Cumple con el objetivo	Iteración en la que se cumplió el objetivo	Referencia hacia la imagen
OBJ-1	Cumple satisfactoriamente	Tercera iteración	Fig. 5
OBJ-2	Cumple satisfactoriamente	Primera iteración	Fig. 6
OBJ-3	Cumple satisfactoriamente	Primera iteración	Fig. 7
OBJ-4	Cumple satisfactoriamente	Segunda iteración	Fig. 8
OBJ-5	Cumple satisfactoriamente	Segunda iteración	Fig. 9
OBJ-6	Cumple satisfactoriamente	Tercera iteración	Fig. 10

#### a) Primera Iteración

En esta etapa se abordó el modelado inicial contemplando variantes, entre ellas la incrustación de la edad; por otro lado, también se filtró el conjunto de datos para procesar los datos de pacientes positivos a COVID-19. Además se comenzó a definir el conjunto de APIs (Application Programming Interface) empleadas, como TensorFlow/ Keras, las cuales podían ayudar en la creación de redes neuronales. Dentro de la primera iteración también se analizó el cambio en el dataset

para los tipos de datos apropiados, según la ingeniería de atributos, definiendo varias columnas normalizadas a valores de 0 y 1. También se normalizaron las edades que fueron añadidas como observación en esta iteración, empleado la normalización lineal. Es importante recalcar que en esta iteración se comenzó a plasmar el OBJ-3 para la implementación de las técnicas, ya que este compartía características con varios de los demás. Entonces se utilizó como base para avanzar en la iteración 1, ya que una vez funcionando las técnicas con un objetivo, se podía replicar el trabajo en otros objetivos. También se avanzó sobre el OBJ-2, en el cual se utilizaron gráficas para poder mostrar los porcentajes descriptivos de cada columna.

### b) Segunda Iteración

En esta iteración se comenzaron a abordar los objetivos restantes que compartían características con los objetivos planteados en la iteración anterior, ya que se habían alcanzado buenas métricas para la implementación de la RL y las RNA. En seguida se procedió a la implementación de los OBJ-4 y OBJ-5, replicando el trabajo que se hizo en el OBJ-3. Para lograr este cometido, se cambiaron las variables dependientes por síntomas severos y fallecimiento respectivamente. En esta iteración se hizo también la inserción de estilos y un dashboard a la página web para agregar usabilidad al usuario final. También se colocó la posibilidad de que el usuario pudiera subir un dataset más actualizado de acuerdo a la información nueva disponible en la página oficial de Gobierno.

### c) Tercera y última iteración.

En esta iteración se comenzó con el objetivo 6 (OBJ-6), el cual tenía como cambio, frente a la implementación de los otros objetivos, el uso de fechas como variables independientes; es decir, que estas variables serían las que entrenarían las técnicas de RL y RNA. Para esto fue necesario ajustar las fechas utilizando normalización lineal. Una vez con las fechas normalizadas, se creó la red neuronal con el objetivo de lograr predecir, con base en la fecha que el que paciente comenzó a desarrollar la sintomatología, y la fecha en la que el paciente fue ingresado a la unidad médica, la probabilidad de no recuperación a la enfermedad. También se pulió la aplicación web, tanto en validaciones como refactorización del código. En estas iteraciones se logró llegar a los objetivos planteados para este artículo de manera satisfactoria, además de que se resolvieron muchos problemas que fueron saliendo durante el diseño y desarrollo. Adicionalmente se creó la herramienta esperada, terminando así con el Objetivo 1 (OBJ-1). Dentro de las iteraciones también se modificaron valores de las redes neuronales para intentar mejorar su precisión, pero no se lograron cambios radicales, aunque si pequeños; al final, las redes neuronales quedaron con la configuración óptima. Terminando en esta iteración con todos los objetivos completados.

### E. Evaluar el funcionamiento del prototipo.

La regresión logística, para la predicción de la probabilidad de desarrollar síntomas severos (intubación), tuvo un 98% de precisión, un 1.0 de recall y un 0.99 de F1-Score. El modelo que predice la probabilidad de fallecimiento tuvo un 92.3% de precisión, un 0.99 de recall y un 0.96 de F1-Score. El último modelo de RL que predice la

probabilidad de que un paciente sea hospitalizado o ambulatorio obtuvo una precisión de 85.6%, un 0.97 de recall y 0.92 de F1-Score. Como se puede observar, la precisión entre los modelos no se alejan mucho uno de otra, con esto se catalogan el uso de estos modelos en este campo como confiables. Mientras que los modelos de redes neuronales también mostraron buenas métricas, la primera red neuronal que predice la probabilidad de desarrollar síntomas graves tuvo una precisión de 98%, un recall de 0.907 y una especificidad de 0.866. La segunda red neuronal que predice la probabilidad de fallecimiento tuvo una precisión de 92.4%, con un recall de 0.957 y una especificidad de 0.929. La tercera red que predice la probabilidad de hospitalización de un paciente tuvo 85.6% de precisión, un recall de 0.902 y una especificidad de 0.904. Por último, la red neuronal que predice el posible fallecimiento del paciente, según la demora que presentó síntomas y la fecha que fue ingresado al hospital, obtuvo una precisión de 92.3%, un recall de 0.774 y una especificidad de 0.745. Notando así que ambos modelos fueron considerablemente eficientes y confiables. Por el lado de la interfaz de usuario, se visualiza que aún existe espacio para mejorar en futuras versiones. En cuanto a las gráficas de los datos en general, también existe espacio de mejora, pero por el momento se considera que es funcional a usuario final.

### 1) Análisis descriptivo de gráficas proporcionadas por la aplicación.

Primero que nada, algo de suma importancia en el sistema web propuesto es la opción de subir un dataset cada vez que se disponga de una versión actualizada del mismo (ver Fig. 5).

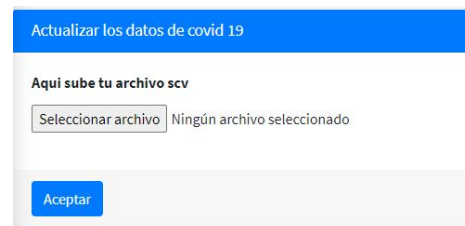


Fig. 5. Opción de subir nuevos archivos csv para actualizar los datos

Una vez ingresado un dataset, es posible obtener información descriptiva de este, por ejemplo, un campo importante es la columna de clasificación final de la prueba COVID-19 aplicada (ver Fig. 6.), en donde se tiene que:

- Caso de COVID-19 confirmado por asociación clínica epidemiológica. Este es confirmado por asociación, aplica cuando el caso informó ser contacto de un positivo COVID-19 y al caso no se le tomó muestra o la muestra resultó no válida. En esta categoría se tiene a un 2.1% del dataset.
- Caso de COVID-19 confirmado por comité de dictaminación. Este es confirmado por dictaminación, solo aplica para las defunciones que cumplan las condiciones de: al caso se le tomó muestra, pero la muestra resultó no válida, o bien no se le tomó muestra. En esta categoría se tienen el 0.1% del dataset.
- Caso de SARS-CoV-2 confirmado. Este aplica cuando: el caso de muestra de laboratorio o prueba





configuraciones de las redes neuronales, además de utilizar diferentes formas de normalización u otras técnicas de Machine Learning. Para futuros trabajos al respecto se recomienda experimentar con diferentes configuraciones de funciones de optimización, tasas de aprendizaje, funciones de activación, épocas de entrenamiento, variables dependientes, además de utilizar diferentes técnicas de Machine Learning para obtener otros resultados.

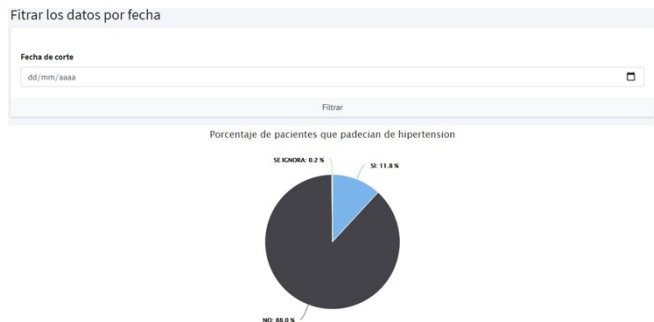


Fig. 10. Gráfica de los porcentajes de los pacientes registrados con hipertensión

Por otro lado, en cuanto a las interfaces web que describen los mecanismos de predicción, en las Fig. 11, 12, 13 y 14 se muestran respectivamente ejemplos de casos en donde se desean predecir tanto posible hospitalización según la edad y enfermedades o factores de riesgo, predecir posible intubación según la edad y enfermedades o factores de riesgo, predecir posible fallecimiento según la edad y enfermedades o factores de riesgo, así como predecir posible fallecimiento según la fecha en la que fue ingresado el paciente a la unidad médica. En estas interfaces el usuario indica las entradas al modelo, para posteriormente obtener el porcentaje de predicción según el tipo de opción solicitado.

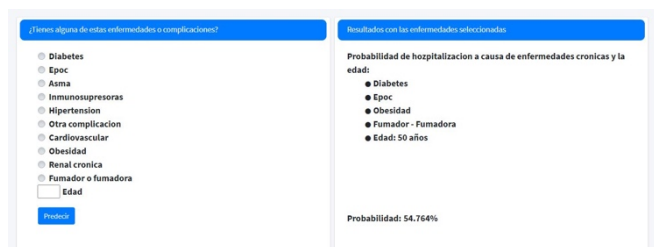


Fig. 11. Opción para predecir posible hospitalización según la edad y enfermedades o factores de riesgo

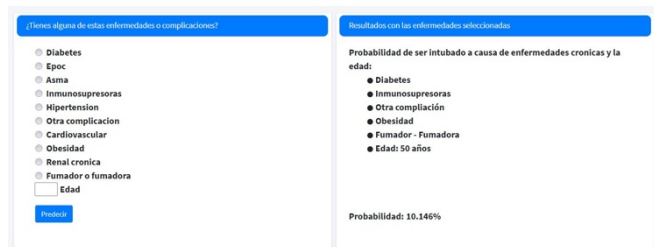


Fig. 12. Opción para predecir posible intubación según la edad y enfermedades o factores de riesgo

Adicionalmente en el sistema web se disponen de otros tipos de análisis descriptivos, como matrices de correlación de variables, datos de lugares geográficos más propensos, información sobre el sexo de las personas analizadas, entre otros.

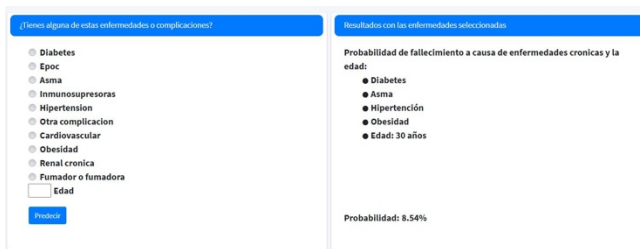


Fig. 13. Opción para predecir posible fallecimiento según la edad y enfermedades o factores de riesgo

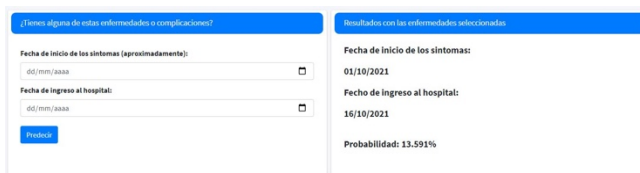


Fig. 14. Opción para predecir posible fallecimiento según la fecha en la que fue ingresado el paciente a la unidad médica

#### IV. DISCUSIÓN

El proceso que se utilizó en la presente investigación fue el que más se acomodó a las necesidades del dataset. Las técnicas de Machine Learning que se utilizaron fueron recomendadas por trabajos apropiados de acuerdo al estado del arte. A pesar de que los resultados de precisión de algunas las redes neuronales no fueron tan altos, se fueron mejorando con pequeños ajustes en los parámetros e hiper-parámetros para poder obtener mejores valores, esto resultando en los porcentajes de precisión finales expuestos en este trabajo. La regresión logística resulto un gran procedimiento de análisis de datos, dadas su altas fiabilidades en precisión. Aún queda camino por recorrer en mejorar las tareas de la presente investigación, sin embargo, algo que cabe la pena resaltar es la capacidad de generar información automática del dataset a partir de la aplicación web, la cual se catalogó como descriptiva e inferencial.

Los principales aportes del presente trabajo se enfocan en brindar una herramienta de software de fácil uso y que sirve como referencia para trabajar con la información del dataset gubernamental, proporcionando la capacidad de subir nuevas versiones y brindar datos actualizados procesados. Es así que el usuario solo tiene que subir un nuevo archivo y trabajar con esos datos, a la vez que si accede a una versión previamente procesada, los estadísticos se verán reflejados con mayor agilidad.

Bajo la base de la funcionalidad del prototipo presentados en esta investigación, se debe tomar en cuenta que cualquier sistema es expuesto a consideraciones de cambios futuros, tanto en el diseño como en el propio esquema del software desarrollado. Se sintetizan los cambios expuestos que quedan abiertos en la propuesta actual; en donde las “interfaces” representan los elementos gráficos (con probabilidad de cambio mediana), la “funcionalidad” implica los requisitos del usuario (con probabilidad de cambio alta), los “datos” (con probabilidad de cambio mediana) y “funciones” (con probabilidad de cambio alta) equivalen a los componentes internos que se usan para describir a los “objetos” (correspondientes a las estructuras de datos básicas de la programación, con probabilidad mediana de cambio), en tanto que la “información” (con probabilidad de cambio alta) representa el dominio del problema en una aplicación. La

arquitectura del sistema debe distinguir entre elementos con mayor y menor probabilidad de cambios. De igual manera, el desarrollo del software debe contemplar un modelo de procesos en el que los componentes de mayor probabilidad de cambio no “arrastren” a los más estables. Por ende, el catálogo de funcionamientos y herramientas proporcionadas por el prototipo presentado aquí propician un cambio constante en su diseño para adecuarse a los momentos y lugares físicos donde se implemente.

#### V. CONCLUSIONES Y TRABAJO FUTURO

El objetivo de esta investigación fue crear un prototipo de sistema web capaz de brindar análisis de información descriptiva e inferencial; en donde se puede predecir, con base en los factores de riesgo de una persona, su probabilidad de ser hospitalizado, la probabilidad de padecer síntomas graves (como la intubación), y la probabilidad de no sobrevivir a la enfermedad. Esta última también utilizando como entrada la fecha en la que comenzó a presentar los síntomas y la fecha en la que fue ingresada a la UM; además de brindar información descriptiva de los datos del COVID-19 que se tienen en México. Todo esto con datos oficiales de GOB de México. Para lograr el cometido, se utilizaron en específico dos técnicas de Machine Learning, que fueron regresión logística y redes neuronales, en las que el proceso de entrenamiento conllevó que los datos fueran tratados y normalizados.

Los principales resultados de las técnicas que se utilizaron mostraron que las redes neuronales obtuvieron una precisión para predicción de hospitalización de 85.6%, para predicción de muerte se tuvo 92.4% y 92.3% de precisión para el modelo que utiliza factores de riesgo y el modelo que utiliza las fechas de síntomas e ingreso a la UM respectivamente; mientras que se obtuvo el 98% de precisión para el modelo que predice la probabilidad de desarrollar síntomas graves. Los resultados de regresión logística obtuvieron una precisión del 92.3% para calcular la posible defunción de un paciente, 98% de precisión para calcular la probabilidad de que un paciente termine siendo intubado debido a su condición de salud; mientras que la predicción de la posible hospitalización de un paciente obtuvo un resultado de 85.6% de precisión. Con esto se concluye que se puede disponer de un sistema de apoyo de análisis de datos para usuarios finales utilizando Machine Learning, brindando información confiable en apoyo a etiquetar a pacientes propensos y poder tomar decisiones tempranas de cómo llevar su cuidado. Para trabajos futuros sería convenientes poder indagar en otro tipo de técnicas de Machine Learning para comparar resultados, además de crear un nuevo prototipo utilizando un dataset que tenga variables para conocer si el paciente fue vacunado, con qué tipo de vacuna, cuántas dosis lleva, entre otros datos. Esto podría brindar mejor información de tendencias y situaciones a considerar en el modelado de casos de salud social, pudiendo indagar el efecto que se tiene en los pacientes que contraen la enfermedad y que a la vez pueden estar ya vacunados.

#### REFERENCIAS

- [1] World Health Organization, “Coronavirus disease (COVID-19) pandemic,” 2021. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019> (accessed Oct. 11, 2021).
- [2] M. Palacios Cruz, E. Santos, M. A. Velázquez Cervantes, and M. León Juárez, “COVID-19, a worldwide public health emergency,” *Revista Clínica Española*, vol. 221, no. 1. Elsevier Doyma, pp. 55–61, Jan. 01, 2021. doi: 10.1016/j.rce.2020.03.001.
- [3] C. Huang *et al.*, “Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China,” *The Lancet*, vol. 395, no. 10223, pp. 497–506, Feb. 2020, doi: 10.1016/S0140-6736(20)30183-5.
- [4] C. del Rio and P. N. Malani, “COVID-19 - New Insights on a Rapidly Changing Epidemic,” *JAMA - Journal of the American Medical Association*, vol. 323, no. 14. American Medical Association, pp. 1339–1340, Apr. 14, 2020. doi: 10.1001/jama.2020.3072.
- [5] W. Zhang *et al.*, “Molecular and serological investigation of 2019-nCoV infected patients: implication of multiple shedding routes,” *Emerging Microbes and Infections*, vol. 9, no. 1, pp. 386–389, Jan. 2020, doi: 10.1080/22221751.2020.1729071.
- [6] X. Escudero, J. Guarner, A. Galindo-Fraga, M. Escudero-Salamanca, M. A. Alcocer-Gamba, and C. Del-Río, “The SARS-CoV-2 (COVID-19) coronavirus pandemic: Current situation and implications for Mexico,” *Archivos de Cardiología de Mexico*, vol. 90, pp. 7–14, May 2020, doi: 10.24875/ACM.M20000064.
- [7] S. Sanche, Y. T. Lin, C. Xu, E. Romero-Severson, N. Hengartner, and R. Ke, “RESEARCH High Contagiousness and Rapid Spread of Severe Acute Respiratory Syndrome Coronavirus 2,” *Emerging Infectious Diseases*, vol. 26, no. 7, pp. 1470–1477, Jul. 2020, doi: 10.3201/eid2607.200282.
- [8] Gobierno de México, “Covid-19 México,” 2021. <https://datos.covid-19.conacyt.mx/> (accessed Oct. 12, 2021).
- [9] World Health Organization, “Situation Report-13,” 2020. Accessed: Oct. 12, 2021. [Online]. Available: <https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200202-sitrep-13-ncov-v3.pdf>
- [10] Gobierno de México, “COVID-19,” 2021. <https://coronavirus.gob.mx/> (accessed Oct. 20, 2021).
- [11] Django, “Django.” <https://www.djangoproject.com/> (accessed Oct. 12, 2021).
- [12] M. Sánchez-Montañés, P. Rodríguez-Belenguer, A. J. Serrano-López, E. Soria-Olivas, and Y. Alakhdar-Mohmara, “Machine learning for mortality analysis in patients with COVID-19,” *International Journal of Environmental Research and Public Health*, vol. 17, no. 22, pp. 1–20, Nov. 2020, doi: 10.3390/ijerph17228386.
- [13] N. S. Punn, S. K. Sonbhadra, and S. Agarwal, “COVID-19 epidemic analysis using machine learning and deep learning algorithms,” *medRxiv*, 2020, doi: 10.1101/2020.04.08.20057679.
- [14] M. Jamshidi *et al.*, “Artificial Intelligence and COVID-19: Deep Learning Approaches for Diagnosis and Treatment,” *IEEE Access*, vol. 8, pp.

- 109581–109595, 2020, doi: 10.1109/ACCESS.2020.3001973.
- [15] X. Li *et al.*, “Deep learning prediction of likelihood of ICU admission and mortality in COVID-19 patients using clinical variables,” *PeerJ*, vol. 8, Nov. 2020, doi: 10.7717/peerj.10337.
- [16] A. Abdulaal, A. Patel, E. Charani, S. Denny, N. Mughal, and L. Moore, “Prognostic modeling of COVID-19 using artificial intelligence in the United Kingdom: Model development and validation,” *Journal of Medical Internet Research*, vol. 22, no. 8, Aug. 2020, doi: 10.2196/20259.
- [17] C. Iwendi *et al.*, “COVID-19 patient health prediction using boosted random forest algorithm,” *Frontiers in Public Health*, vol. 8, Jul. 2020, doi: 10.3389/fpubh.2020.00357.
- [18] S. Lewis, “What is Prototyping Model?,” 2005. <https://searchcio.techtarget.com/definition/Prototyping-Model> (accessed Oct. 20, 2021).
- [19] A. Weitzenfeld and S. Guardati, “Capítulo 12 Ingeniería de software: el proceso para el desarrollo de software.”
- [20] I. Sommerville, *Engineering software products*. Boston: Addison Wesley, 2020.
- [21] A. Becerra *et al.*, “Learning content management software personalized for a university environment”, vol. 40, no. 130, pp. 347-362.
- [22] Gobierno de México, “Datos de gobierno,” 2021. <https://datos.gob.mx/> (accessed Oct. 20, 2021).
- [23] H. Albino, D. Guillermo, and G. Murillo, “Libro-Redes neuronales Related papers Redes Neuronales Artificiales y sus Aplicaciones Ronald Mora REDES NEURONALES ARTIFICIALES Y SUS APLICACIONES Tabla de Contenidos.”
- [24] R. N. Singarimbun, E. B. Nababan, and O. S. Sitompul, “Adaptive Moment Estimation to Minimize Square Error in Backpropagation Algorithm,” Nov. 2019. doi: 10.1109/ICoSNIKOM48755.2019.9111563.
- [25] R. O. Sinnott, H. Duan, and Y. Sun, “A Case Study in Big Data Analytics: Exploring Twitter Sentiment Analysis and the Weather,” in *Big Data: Principles and Paradigms*, Elsevier Inc., 2016, pp. 357–388. doi: 10.1016/B978-0-12-805394-2.00015-5.
- [26] S. García, J. Luengo, and F. Herrera, “Intelligent Systems Reference Library 72 Data Preprocessing in Data Mining.” [Online]. Available: <http://www.springer.com/series/8578>
- [27] S. Raschka and V. Mirjalili, *Python machine learning: machine learning and deep learning with python, scikit-learn, and tensorflow 2*. 2021.